2-3-2007

# The Relationship Between Discrete Vector Quantization and the P-Median Problem

Allen G. Holder
*Trinity University*, aholder@trinity.edu

G Lim

J Reese

# The Relationship Between Discrete Vector Quantization and the P-Median Problem

A. Holder[†*], G. Lim[‡], and J. Reese[‡]

February 3, 2007

**Abstract**

We show that a well studied problem in the engineering community is the same as a problem studied by mathematical combinatorialists. Specifically, we show that the question of optimally designing a vector quantizer, which is an important problem in coding theory, is the same as the $p$-median problem, which is a classic graph theory problem with important applications in operations research. The importance of the relationship lies in the fact that both communities have spent years developing solution methodologies, and this connection permits each community to glean new ideas from the other. We show that two of the most popular heuristics are equivalent, meaning that they produce the same sequence of iterates, under suitable conditions. However, the technique used to design a quantizer outperforms its counterpart in graph theory under these same conditions.

**Keywords:** Vector Quantization, P-Median problem, Optimization, Facility Location, Lloyd Algorithm, Maranzana Algorithm.

---

† Trinity University Mathematics, San Antonio, TX USA.

‡ University of Houston, Department of Industrial Engineering, Houston, TX USA.

* Corresponding author - aholder@trinity.edu

# 1 Introduction

Advances in information and coding theory have revolutionized how society handles data transactions. The examples are numerous and far reaching and include internet transmissions, image compression, and security coding, to name just a few. Information theory is credited to Claude E. Shannon, and one of the problems he foreshadowed was that of vector quantization (VQ). Shannon's insight was that we can quantify the difference between an input and coded signal by assuming the existence of a distortion measure. Although not titled vector quantization at the time, he worked on VQ under the guise of *source coding subject to a fidelity criterion* [3].

About the same time as Shannon was laying the groundwork for information theory, graph theorists and operations researchers were beginning to investigate a clustering problem called the $p$-median problem. Initial investigations were undertaken by Hakimi [6], and investigations have continued as part of facility location. The problem has spawned a substantial literature, and readers are directed to the annotated bibliography in [11]. At first glance, there is no reason to believe that coding theory should have any but the most tenuous connection to facility location. However, both problems are clustering problems since both are interested in grouping elements and assigning a representative to each group. So there surely is a connection, and in fact, we show that the problems in the discrete setting are equivalent. Both problems have continuous counterparts, with VQ's original intent being to quantize a continuous signal and the $p$-median's foundations lying in the Weber problem [11, 18]. As with all continuous problems that are solved with digital computation, both manifest themselves as discrete problems in practice. Moreover, in the case of the $p$-median problem Hakimi [6] proved that the continuous problem always has a solution that corresponds to its discrete counterpart.

We begin by defining the problems in detail and establishing the previously unrecognized result that they are the same. The sense of equivalence is strong, and in fact, a primal and dual pair of linear programs would not be equivalent with our definition. The connection is important for two reasons: 1) some problems are conceptually easier to model in terms of quantizer design while others are better suited to graph theory descriptions, and 2) each community can adopt the heuristic procedures developed in the other community. With regards to the latter, we show that under suitable conditions two of the most popular heuristics produce the same sequence of iterates, but that the technique developed for VQ design is superior computationally.

# 2 Notation and Problem Statements

Vector quantization is a process that codes continuous or discrete signals subject to a fidelity criterion and is often used to compress images or other data. A *vector quantizer*, or simply a quantizer, is a mapping $\mathcal{Q}$ from an input set of vectors $\mathbb{V}$ onto an $N$ element subset $\mathcal{C}$ of $\mathbb{V}$ —i.e. $\mathcal{Q} : \mathbb{V} \overset{\text{onto}}{\to} \mathcal{C} \subseteq \mathbb{V}$, where $|\mathcal{C}| = N$. The image set $\mathcal{C}$ is called the *codebook* and its elements are called *codevectors* or *codewords*. A quantizer partitions $\mathbb{V}$ into $N$ distinct regions called *cells*, which are defined for every $\hat{v} \in \mathcal{C}$ by

$$\mathbb{V}_{\hat{v}} = \{v \in \mathbb{V} \mid \mathcal{Q}(v) = \hat{v}\}.$$

Quantizers are typically separated into two processes, known as the *encoder* $\mathcal{E}$ and the *decoder* $\mathcal{D}$. The encoder's task is to assign an input vector to a partition cell, and hence, $\mathcal{E} : \mathbb{V} \to \{1, 2, \ldots, N\}$. The decoder selects a vector from each cell to serve as that cell's codevector. So, $\mathcal{D} : \{1, 2, \ldots, N\} \to \mathcal{C}$ and $Q(v) = \mathcal{D}(\mathcal{E}(v))$. A common example is found in analog to digital conversion, where the continuous analog signal in Hertz (Hz) is quantized into a finite collection of digital signals. For example, a human's auditory range is between 20 and 20,000 Hz. If the digital storage medium only distinguishes between $N$ different signals, a simple encoder would map the interval $[20{+}19980(i{-}1)/N\,,\ 20{+}19980i/N]$ to the integer $i$, for $i = 1, 2, \ldots, N$. A simple decoder would map $i$ to the midpoint of the interval, $i \mapsto 20{+}19980(i{-}1/2)/N$. Such a quantizer mimics the rounding process. A discrete example is to let $\mathbb{V}$ be a finite set of points on a city map. Consider an encoder that maps these locations into school districts, each represented by a district number. The decoder then maps the index number of a particular district to the location for that district's school. In this work we only consider discrete quantizers, and hence $|\mathbb{V}| \leq |\mathbb{N}|$, where $\mathbb{N}$ is the set of natural numbers.

A quantizer's performance is evaluated in terms of *distortion*, which relies on two pieces of information. The first of these is a real value that represents the similarity between any two input vectors, and allowing $\mathbb{R}_+$ to be the set of nonnegative real values, we let $\rho : \mathbb{V} \times \mathbb{V} \to \mathbb{R}_+$ map the ordered pair $(v_i, v_j)$ to the nonnegative similarity $\rho(v_i, v_j)$, which measures how similar $v_i$ is to $v_j$. We emphasize that $\rho$ is not necessarily a metric. In particular, $\rho$ is not required to be symmetric, it does not need to satisfy the triangular inequality, and $\rho(v_i, v_j) = 0$ does not necessarily imply $v_i = v_j$. This said, a common measure is the squared Euclidean distance, $\rho(v_i, v_j) = \|v_i - v_j\|^2$, but the particular measure is often tailored to the application. We simply assume that $\rho(v_i, v_j)$ is a nonnegative value meaningful to its application. The second piece of information is the probability of observing an input vector, and we assume that $\alpha(v)$ is the probability of observing $v$. In the school district example, $\alpha(v)$ could be the probability that location $v$ is a house with school-age children.

A quantizer's performance is measured in terms of its expected (or average) distortion, which is

$$D_{\mathcal{Q}} = E\rho(v, \mathcal{Q}(v)) = \sum_i \rho(v_i, \mathcal{Q}(v_i))\alpha(v_i),$$

where $i$ indexes the elements of $\mathbb{V}$. The objective is to design an optimal quantizer, meaning that it minimizes the expected distortion. The feasible region for the design process is the collection of $N$ element subsets of $\mathbb{V}$, making the size of the feasible region $\binom{|\mathbb{V}|}{N}$. For any particular $N$ element subset of $\mathbb{V}$, say $\mathbb{W}$, we have that the partition cell for each $w \in \mathbb{W}$ that minimize the distortion is

$$\mathbb{V}_w = \{v \in \mathbb{V} : \rho(v, w) \leq \rho(v, w') \text{ for } w' \in \mathbb{W}\}. \tag{2.1}$$

A subtle nuance is that $\mathbb{V}_w$ is not necessarily unique because some of the elements of $\mathbb{V}$ may be equally similar to several elements of $\mathbb{W}$. In such a situation, we assign these elements of $\mathbb{V}$ to the member of $\mathbb{W}$ with the smallest index, making $\mathbb{V}_w$ well defined. With this notation, the *$N$-element VQ design problem on $\mathbb{V}$ with respect to* $(\rho, \alpha)$ is

$$\min \left\{ \sum_{w \in \mathbb{W}} \sum_{v \in \mathbb{V}_w} \rho(v, w)\alpha(v) : \mathbb{W} \subseteq \mathbb{V},\ |\mathbb{W}| = N \right\}. \tag{2.2}$$

We now turn our direction to the $p$-median problem, which is a facility location problem first investigated by Hakimi [6]. This problem operates on a strongly connected digraph $(\mathbb{V}, \mathbb{E})$. We define a *position* on $(\mathbb{V}, \mathbb{E})$ to be either a vertex or a point along an arc, and we write $\mathbb{P} \subseteq (\mathbb{V}, \mathbb{E})$ to mean that $\mathbb{P}$ is a collection of positions on $(\mathbb{V}, \mathbb{E})$. Although it is customary to call the problem the $p$-median problem, for notational convenience we instead refer to it as the $N$-median problem ($p$ is used denote a position). We let $\beta(v)$ be the weight assigned to vertex $v$ and $\gamma(p_i, p_j)$ be the nonnegative value assigned to each ordered pair of positions $(p_i, p_j)$. As before, we do not generally assume that $\gamma$ is a metric. The $N$-median problem is to locate $N$ positions on the digraph and an assignment of the vertices to these positions that solves

$$\min \left\{ \sum_{p \in \mathbb{P}} \sum_{v \in \mathbb{V}_p} \gamma(p, v)\beta(v) : \mathbb{P} \subseteq (\mathbb{V}, \mathbb{E}),\ |\mathbb{P}| = N \right\},$$

where

$$\mathbb{V}_p = \{v \in \mathbb{V} : v \text{ is assigned to position } p\}.$$

Any collection $\mathbb{P}$ that solves this optimization problem is said to be a collection of *medians*.

While the vertex set is assumed to be finite, the fact that the positions are allowed to be located along the edges makes the problem continuous. We are only concerned with the discrete counterpart, which considers selecting $N$ positions from a set $\mathbb{P}'$ with the properties that $\mathbb{V} \subseteq \mathbb{P}'$ and $|\mathbb{P}'| \leq |\mathbb{N}|$. We call this *the $N$-median problem on* $(\mathbb{V}, \mathbb{E})$ *restricted to* $\mathbb{P}'$ *with respect to* $(\gamma, \beta)$. Similar to the quantization problem, for an $N$ element sub-collection $\mathbb{P}$ of $\mathbb{P}'$ to be a collection of medians, it is necessary that

$$\mathbb{V}_p = \{v \in \mathbb{V} : \gamma(v, p) \leq \gamma(v, p') \text{ for } p' \in \mathbb{P}\}. \tag{2.3}$$

Again, these sets are not necessarily unique, and if ties exist, the vertex is assigned to the position with the lowest index. The $N$-median problem restricted to $\mathbb{P}'$ with respect to $(\gamma, \beta)$ is

$$\min \left\{ \sum_{p \in \mathbb{P}} \sum_{v \in \mathbb{V}_p} \gamma(v,p)\beta(v) : \mathbb{P} \subseteq \mathbb{P}', \ |\mathbb{P}| = N \right\}. \tag{2.4}$$

Some historical notes about the $N$-median problem are worth mentioning because they position our work within the current research environment. Hakimi's original work [6] assumed a (strongly) connected and non-negatively weighted (di)graph. From this graph he constructed a complete (di)graph in which the edge weights were the shortest path distances in the original graph. This construction induces the shortest-path metric on the complete graph, which is a pseudo-metric instead of a metric. Under these conditions he established the following result.

**Theorem 2.1 (Hakimi [6])** *If $G$ is a connected (di)graph with nonnegative vertex and edge weights, then there is a collection of $N$ vertices that are also medians with respect to the shortest-path-metric.*

This result states that we can solve the $N$-median problem by restricting our search to vertices, provided that the conditions of the theorem are satisfied. However, over time the conditions guaranteeing this result were disregarded, and today the $N$-median problem is often stated as the discrete problem of finding $N$-vertices on a weighted (di)graph where the edge weights have no particular structure. This modern re-formulation undermines Hakimi's original work, which provides conditions under which a continuous problem can be solved by a discrete counterpart. In the next section, we define an equivalence relation between optimization problems and show that the discrete version of the $N$-median problem is equivalent to finding $N$ vertices of a uniquely associated digraph. This result addresses the modern re-formulation since we do not require any particular structure on the edge weights, and hence, our result re-casts Hakimi's original work with respect to the modern statement of the $N$-median problem. The same equivalence relation is used to show that the optimal design of a quantizer is the same as solving the $N$-median problem. The problem statements in (2.2) and (2.4) certainly hint at this connection, which is formalized in the next section.

# 3 Problem Equivalence

A standard concept of 'equivalent' mathematical programs is not widely accepted within the optimization community. Some use the term to mean that the solutions of one problem correspond to the solutions of another. Others use the term to mean that a problem is simply re-modeled. In fact, nothing is mentioned about this in the *Mathematical Programming Glossary* [4] because the concept is not generally well defined [5]. We define an equivalence relation that requires problems to be invertible transformations of each other, and such problems are called *identical*. To be precise, we say that the problems

$$P1 : \min\{f(x) : x \in X\} \text{ and } P2 : \min\{g(y) : y \in Y\} \tag{3.1}$$

are identical under $h$ if there is a bijection $h : X \to Y$ such that $f = g \circ h^{-1}$. This sense of equivalence is strong and essentially states that we have simply re-labeled the elements of the feasible region in a way that maintains the objective value. Some immediate observations about this relation are that

1. if $P1$ is identical to $P2$ under $h$, then

$$h(\text{argmin}\{f(x) : x \in X\}) = \text{argmin}\{g(x) : y \in Y\}) \text{ and}$$

2. the equivalence class of $P1$, denoted $[f, X]$, is

$$[f, X] = \{(g, Y) : h(X) = Y \text{ and } f = g\circ h^{-1}, \text{ for some bijection } h : X \to Y\}.$$

The problem statements in (2.2) and (2.4) were modeled to highlight their similarity. Indeed, we purposefully used $\mathbb{V}$ to denote both the set of vectors to be quantized and the vertex set of the digraph to highlight this connection. The electrical engineers were addressing the problem probabilistically as the optimal design of a quantizer (recall that $\alpha$ is a probability density) whereas the mathematical programmers approached it as a combinatorial optimization problem. In the end, the problems are identical, a statement made rigorous in Theorem 3.1.

**Theorem 3.1** *Let $\mathbb{P}'$ be a discrete collection of positions on the strongly connected digraph $(\mathbb{V}, \mathbb{E})$. Let $\beta : \mathbb{P}' \to \mathbb{R}_+$ satisfy $\sum_{v\in\mathbb{V}} \beta(v) = 1$ and $\beta(v) = 0$ if $v \in \mathbb{P}'\backslash\mathbb{V}$. Further assume that $\gamma$ is any map from $\mathbb{P} \times \mathbb{P}$ into $\mathbb{R}_+$. Then the following problems are identical.*

1. *The $N$-element VQ design problem on $\mathbb{P}'$ with respect to $(\gamma, \beta)$,*

2. *The $N$-median problem on the complete digraph $(\mathbb{P}', \mathbb{P}' \times \mathbb{P}')$ restricted to $\mathbb{P}'$ with respect to $(\gamma, \beta)$, and*

3. *The $N$-median problem on $(\mathbb{V}, \mathbb{E})$ restricted to $\mathbb{P}'$ with respect to $(\gamma, \beta)$.*

**Proof:** From (2.2) and (2.4) we see that problems 1 and 2 are respectively

$$\min\left\{\sum_{w\in\mathbb{W}}\sum_{v\in\mathbb{V}_w} \gamma(v, w)\beta(v) : \mathbb{W} \subseteq \mathbb{P}', \ |\mathbb{W}| = N\right\} \tag{3.2}$$

and

$$\min\left\{\sum_{p\in\mathbb{P}}\sum_{v\in\mathbb{V}_p}\gamma(v,p)\beta(v) : \mathbb{P}\subseteq\mathbb{P}',\ |\mathbb{P}|=N\right\}. \tag{3.3}$$

We define the bijection from the feasible region of (3.2) onto (3.3) by

$$h(\mathbb{W})=\mathbb{P} \text{ if and only if } \mathbb{W}=\mathbb{P}.$$

We mention that this is nothing more than the identity map on the collection of $N$-element subsets of $\mathbb{P}'$. To see that the objective values align properly under this bijection, notice that from (2.1) and (2.3) we have

$$\begin{aligned}\mathbb{V}_w &= \{v\in\mathbb{P}' : \gamma(v,w)\le\gamma(v,w') \text{ for } w'\in\mathbb{W}\}\\ &= \{v\in\mathbb{P}' : \gamma(v,w)\le\gamma(v,w') \text{ for } w'\in h(\mathbb{W})=\mathbb{P}=\mathbb{W}\}\\ &= \mathbb{V}_p.\end{aligned}$$

Hence, the index sets for the summations agree, from which we conclude that problems 1 and 2 are identical.

To see that problem 3 is identical to problem 2, first notice that problem 3 is (3.3), which means both problems have the same feasible region. This allows us to use the identity map $\hat{h}(\mathbb{P})=\mathbb{P}$. Unfortunately, the index sets of the inner summations do not agree since the vertex set for problem 3 is $\mathbb{V}$ and the vertex set for problem 2 is $\mathbb{P}'$. This means the index set of the inner summation for problem 3 is

$$\mathbb{V}_p=\{v\in\mathbb{V} : \gamma(v,p)\le\gamma(v,p') \text{ for } p'\in\mathbb{P}\},$$

while the index set for problem 2 is

$$\hat{\mathbb{V}}_p=\{v\in\mathbb{P}' : \gamma(v,p)\le\gamma(v,p') \text{ for } p'\in\mathbb{P}\}.$$

However, from the assumption that $\beta(v)=0$ for $v\in\mathbb{P}'\backslash\mathbb{V}$, we have for any $p\in\mathbb{P}'$ that

$$\sum_{v\in\mathbb{V}_p}\gamma(v,p)\beta(v)=\sum_{v\in\hat{\mathbb{V}}_p}\gamma(v,p)\beta(v),$$

and hence the objective values agree under $\hat{h}$. We conclude that problems 2 and 3 are identical. The fact that problems 1 and 3 are identical follows by considering the composition of $h$ and $\hat{h}$. ∎

We mention that the manner in which the two problems were stated is important to the theorem's conclusion. Typically, both problems are stated in terms of selecting a subset of vectors or vertices, referred to as *selection*, and assigning the vectors or vertices to the selected elements, referred to as *assignment*. The optimization problems in (2.2) and (2.4) do not consider assignment in their descriptions of the feasible regions. Instead, both feasible regions are the $N$-element subsets of the vectors or vertices and the assignments are described by the index set of the inner summation. This is allowed because each $N$-element subset defines a unique optimal assignment as defined in (2.1) and (2.3). The fact that some elements may have equal similarity means there may be numerous, even an uncountable number of, alternative assignments. However, the discrete assumption means that the vectors to be quantized and the vertices of the digraph were at most countable. This is crucial to the proof since it allows us to define a unique assignment for each feasible subset with the least index rule.

To highlight the importance of the least index rule, let $(\mathbb{V}, \mathbb{E})$ be the digraph in Figure 1 for problem 3 in Theorem 3.1. The corresponding complete digraph for problem 2 is in Figure 2 (arrows are not shown). The problem on $(\mathbb{V}, \mathbb{E})$ only assigns positions $p_1$, $p_3$ and $p_4$ —i.e. the vertices in $\mathbb{V}$, whereas the problem on $(\mathbb{P}', \mathbb{P}' \times \mathbb{P}')$ assigns $p_1, p_2, \ldots, p_6$ —i.e. the elements of $\mathbb{P}'$. Suppose we are solving the 2-median problem and that $\gamma$ and $\beta$ are such that

- $\{p_1, p_4\}$ is the unique solution (notice this is for both problems), and
- for the problem on $(\mathbb{V}, \mathbb{E})$ we have $\mathbb{V}_{p_1} = \{p_1\}$ and $\mathbb{V}_{p_4} = \{p_3, p_4\}$.

Considering the problem for $(\mathbb{P}', \mathbb{P}' \times \mathbb{P}')$, we see that the positions $p_2$, $p_5$ and $p_6$ now need to be added to $\mathbb{V}_{p_1}$ and $\mathbb{V}_{p_4}$, and our construction says they will be assigned to $p_1$ or $p_4$ depending on to which they are more similar. However, we could have $\gamma(p_i, p_j) = 0$ for $i = 2, 5, 6$ and $j = 1, 3, 5$, meaning $p_2$, $p_5$ and $p_6$ are equally similar to each of the elements in $\mathbb{V}$. Our construction dictates that $p_2$, $p_5$ and $p_6$ are each assigned to the median with the lowest index, and using the notation from the proof of Theorem 3.1, we have $\hat{\mathbb{V}}_{p_1} = \{p_1, p_2, p_5, p_6\}$ and $\hat{\mathbb{V}}_{p_4} = \{p_3, p_4\}$, which is a unique assignment. If the least index rule was removed, then there would be

$$\binom{3}{3} + \binom{3}{2} + \binom{3}{1} + \binom{3}{0} = 8$$

possible ways to add $p_2$, $p_5$ and $p_6$ to $\mathbb{V}_{p_1}$ and $\mathbb{V}_{p_4}$. Notice that if the feasible regions had been stated in terms of both selection and assignment without regard to some tie braking rule for the assignment decision, then this would have violated the fact that the argument minimums would have needed to have the same cardinality. So our statement of the model is important because the objective function is defined in terms of a unique assignment. It is likely that one could address the problem in the continuum by invoking the axiom of choice.

Theorem 3.1 has two important corollaries.

**Corollary 3.2** *Every discrete $N$-element VQ design problem corresponds to a discrete $N$-median problem on a complete digraph restricted the vertices.*

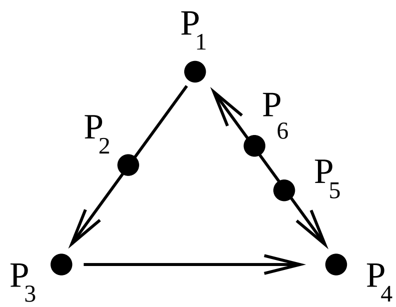


Figure 1: A strongly connected digraph with 3 added positions to select medians from.

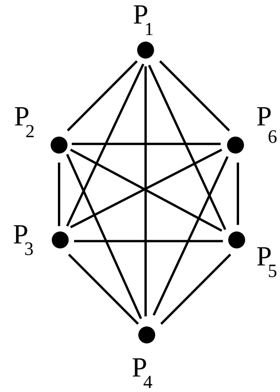


Figure 2: The corresponding complete graph of the digraph in Figure 1, all edges are bi-directional arcs.

This follows immediately from the fact that problems 1 and 2 in Theorem 3.1 are identical. The idea is to start with an $N$-element VQ design problem and simply construct the complete digraph whose vertices are the vectors in the quantization problem. Notice that the $N$-median problem is restricted to the vertices, which allows the similarity measure $\rho$ in the quantization problem to fulfill the role of $\gamma$ in the $N$-median problem. Similarly, the vertex weights in the $N$-median problem are the probability measures in the quantization problem (after normalization).

The second corollary addresses the historical divide between Hakimi's original work and the modern interpretation of the $N$-median problem.

**Corollary 3.3** *For every discrete $N$-median problem there is an alternative $N$-median problem in which only collections of vertices need to be considered. Moreover, this alternative $N$-median problem corresponds to an $N$-element VQ design problem.*

Similar to the previous corollary, this statement follows immediately from the fact that problems 1, 2 and 3 in Theorem 3.1 are identical. However, a graph description is warranted. Consider the $N$-median problem on $(\mathbb{V}, \mathbb{E})$ restricted to $\mathbb{P}'$ with respect to $(\gamma, \beta)$. Recall that $\gamma$ is defined for every element in $\mathbb{P}' \times \mathbb{P}'$. This means we can consider the complete digraph $(\mathbb{P}', \mathbb{P}' \times \mathbb{P}')$ with edge weights defined by $\gamma$. This complete digraph does not have node values for the vertices in $\mathbb{P}' \backslash \mathbb{V}$, and we extend the definition of $\beta$ to $\hat{\beta}$ so that $\hat{\beta}(p) = \beta(p)$ if $p \in \mathbb{V}$ and $\hat{\beta}(p) = 0$ if $p \in \mathbb{P}' \backslash \mathbb{V}$. This extension satisfies the conditions of Theorem 3.1, and hence, we only need to consider collections of vertices of the complete digraph to solve the discrete problem on $(\mathbb{V}, \mathbb{E})$ restricted to $\mathbb{P}'$.

This argument is similar to Hakimi's original proof since he constructs a complete graph with the shortest-path metric. However, Hakimi's proof required an additive property along arcs, namely that if $p$ is a position on $(v_i, v_j)$, then $\gamma(v_i, v_j) = \gamma(v_i, p) + \gamma(p, v_j)$. This is a reasonable assumption in some settings, but it is not appropriate in others. For example, if $\gamma(v_i, v_j)$ is the travel cost from $v_i$ to $v_j$, then stopping along the way may increase or decrease the cost depending on tolls and/or fuel costs. Our approach does not require this additive property. Our conclusion coincides with Hakimi's because they both state that we only need to consider vertices of an appropriate digraph when searching for medians.

# 4 Solution Techniques & Complexity

Kariv and Hakimi [8] showed that the problem of finding an $N$-median on a connected digraph is NP-hard in $N$ and $|\mathbb{V}|$. Polynomial algorithms exist, however, for the case when $N$ is fixed. This is often confused in the literature, where many have supported the need for heuristics because of the NP-hardness of the problem for a fixed $N$. A complete discussion of this topic is found in [2].

Although our discrete version of the $N$-median problem is different than Hakimi's original statement, the following result states the related conclusion that the discrete $N$-median problem on $(\mathbb{V}, \mathbb{E})$ restricted to $\mathbb{P}'$ with respect to $(\gamma, \beta)$ is polynomial as long as $N$ is fixed.

**Theorem 4.1** *The worst-case complexity of the discrete $N$-median problem on $G = (\mathbb{V}, \mathbb{E})$ restricted to $\mathbb{P}'$ with respect to $(\gamma, \beta)$ is $O(|\mathbb{V}||\mathbb{P}'|^{N+1})$.*

**Proof:** The size of the feasible region is $\binom{|\mathbb{P}'|}{N} = O(|\mathbb{P}'|^N)$. We need to compare each element of a feasible $\mathbb{P}$ to the elements of $\mathbb{V}$ to form $\mathbb{V}_p$, which is $O(|\mathbb{V}||\mathbb{P}'|)$. So, starting with $(\mathbb{V}, \mathbb{E})$ and $\mathbb{P}'$, we require no more than $O(|\mathbb{V}||\mathbb{P}'|^{N+1})$ iterations to define (2.4). The addition in the objective function additionally requires no more than $O(N|\mathbb{V}|)$ multiplications. Hence, the total computation requires no worse than $O(|\mathbb{V}||\mathbb{P}'|^{N+1} + N|\mathbb{V}|) = O(|\mathbb{V}||\mathbb{P}'|^{N+1})$. ■

Notice that since $\mathbb{V} \subseteq \mathbb{P}'$, we also find that the complexity is no worse than $O(|\mathbb{P}'|^{N+2})$, which is less impressive for the $N$-median problem but appropriate for VQ design. This leads to the following corollary.

**Corollary 4.2** *The worst-case complexity of the $N$-element VQ design problem on $\mathbb{P}'$ with respect to $(\gamma, \beta)$ is $O(|\mathbb{P}'|^{N+2})$.*

**Proof:** From Theorem 3.1 we have that if $\mathbb{V} = \mathbb{P}'$, then the problems are identical. So the result is established by Theorem 4.1. ■

Polynomial time does not mean heuristics are unimportant, and the computational demand in many applications exceeds modern capabilities. Both communities have suggested heuristics, and a benefit of Theorem 3.1 is that it allows us to model a situation as either VQ design or median location, whichever is cognitively simpler, but heuristically solve the problem with the techniques from the other realm. The rest of this section compares some of the common heuristics for both problems, with the most significant result being that two of the heuristics are the same under common conditions, meaning that they produce the same sequence of iterates. However, a subsequent argument shows that the method for VQ design has a lower complexity.

The most common method for designing a quantizer is the Lloyd Algorithm, which was originally proposed in an unpublished technical report in 1957 and later published in 1982 [9]. Lloyd's algorithm is stated in terms of the continuum, and we discuss and analyze a discrete counterpart that we call the discrete Lloyd algorithm (DLA). With respect to the $N$-median problem, a technique first proposed in 1964 that remains popular is the Maranzana algorithm [10]. Both techniques iterate between the assignment and selection parts of the problem in a way that improves the objective function. In terms of (2.2) and (2.4), both algorithms begin with an initial feasible element, $\mathbb{W}$ and $\mathbb{P}$. The assignments part is the construction of the inner summations' index sets, $\mathbb{V}_w$ for $w \in \mathbb{W}$ and $\mathbb{V}_p$ for $p \in \mathbb{P}$. The selection part of the problem is to update $\mathbb{W}$ and $\mathbb{P}$ by respectively calculating for each $w \in \mathbb{W}$ and $p \in \mathbb{P}$

$$\operatorname{argmin}\left\{\sum_{v \in \mathbb{W}_w} \rho(v, w')\alpha(v) : w' \in \mathbb{W}_w\right\} \tag{4.1}$$

and

$$\operatorname{argmin}\left\{\sum_{v \in \mathbb{V}_p} \gamma(v, p')\beta(v) : p' \in \mathbb{V}_p\right\}. \tag{4.2}$$

An element from each argument minimum is selected to form the new feasible sets, say $\hat{\mathbb{W}}$ and $\hat{\mathbb{P}}$, which replace $\mathbb{W}$ and $\mathbb{P}$. The process continues until $\hat{\mathbb{W}} = \mathbb{W}$ and $\hat{\mathbb{P}} = \mathbb{P}$. The objective function is non-increasing with every new $\hat{\mathbb{W}}$ and $\hat{\mathbb{P}}$, see [3] and [10].

The fact that these two algorithms produce the same iterates if applied to the same problem and initialized in the same way is clear. However, because the historical developments are different, the Lloyd algorithm calculates the argument minimums in (4.1) differently than the Maranzana algorithm does for the argument minimums in (4.2). The difference lies in the fact that VQ design was traditionally addressed as a continuous problem with $\rho(v_i, v_j) = \|v_i - v_j\|^2$ —i.e. the similarity measure between two vectors was the squared error. Lloyd used this to his advantage when calculating the argument minimums, and instead of addressing (4.1) directly, he instead calculated the center-of-mass of each cell $\mathbb{W}_w$, which is

$$\frac{\sum\limits_{v\in\mathbb{W}_w} \alpha(v)v}{\sum\limits_{v\in\mathbb{W}_w} \alpha(v)}.$$

A couple of observations are warranted. This calculation requires the product $\alpha(v)v$ to be well defined, which is true if $\mathbb{V}$ is a vector space built on a scalar field containing the range of $\alpha$. Traditional VQ problems are cast in the continuum with $\mathbb{V} = \mathbb{R}^n$ and $\alpha(\mathbb{V}) \subseteq \mathbb{R}$, an assumption used by Lloyd. Our discrete setting does not permit this assumption, but the dense approximation $\mathbb{V} = \mathbb{Q}^n$ and $\alpha(\mathbb{V}) \subseteq \mathbb{Q}$ is allowed ($\mathbb{Q}$ is the set of rationals), which reduces everything to rational arithmetic. In general we simply assume in this discussion that $\mathbb{V}$ is at most a countable subset of $\mathbb{R}^n$ ($\alpha$ is already assumed to be real valued).

A second observation in the discrete setting is that we are not guaranteed $\mathbb{V}$ contains the center-of-mass even if the arithmetic is well defined. This is not an issue for rational arithmetic but is a significant problem in the common situation of $\mathbb{V}$ being finite. However, if we assume that $\alpha(v)$ is constant, which is the same as assuming the elements of $\mathbb{V}$ are uniformly distributed, then we can compute the center-of-mass and project it onto $\mathbb{W}_w$ to calculate the argument minimums in (4.1). A short argument establishing this fact begins with the well-known and easily established result that

$$\left\{ \frac{1}{|\mathbb{W}_w|} \sum_{v\in\mathbb{W}_w} v \right\} = \operatorname{argmin} \left\{ \sum_{v\in\mathbb{W}_w} \|x - v\|^2 : x \in \mathbb{R}^n \right\}, \tag{4.3}$$

which states that the center-of-mass minimizes the squared error in the continuum. Since the objective is strictly convex, a proof of this follows by showing that the center-of-mass satisfies the first order conditions, an argument we forgo. We let $\text{proj}_{\mathbb{W}_w}(v)$ be the nearest element of $\mathbb{W}_w$ to $v$, with ties being decided by the least index rule, and show that this element is in the desired argument minimum. Similar ideas are found in [1].

**Theorem 4.3** *Assume* $\mathbb{W} \subseteq \mathbb{V}$, $|\mathbb{V}| = N < \infty$, $\rho(v_i, v_j) = \|v_i - v_j\|^2$ *and* $\alpha(v)$ *is constant. Then, for each* $w \in \mathbb{W}$ *we have*

$$\mathit{proj}_{\mathbb{W}_w} \left( \frac{1}{|\mathbb{W}_w|} \sum_{w\in\mathbb{W}_w} w \right) \in \operatorname{argmin} \left\{ \sum_{v\in\mathbb{W}_w} \|w' - v\|^2 : w' \in \mathbb{W}_w \right\}.$$

**Proof:** Let $w \in \mathbb{W}_w$ and $M = (1/|\mathbb{W}_w|) \sum_{w\in\mathbb{W}_w} w$. From (4.3) we have for any nonnegative $k$ that

$$\begin{aligned}
&\left\{x \in \mathbb{R}^n : \sum_{v \in \mathbb{W}_w} \|x - v\|^2 \le \sum_{v \in \mathbb{W}_w} \|M - v\|^2 + k\right\} \\
= &\left\{x \in \mathbb{R}^n : \sum_{v \in \mathbb{W}_w} \big((x-v)^T(x-v) - (M-v)^T(M-v)\big) \le k\right\} \\
= &\left\{x \in \mathbb{R}^n : \sum_{v \in \mathbb{W}_w} \big(x^T x - 2x^T v - M^T M + 2M^T v\big) \le k\right\} \\
= &\Bigg\{x \in \mathbb{R}^n : \frac{1}{|\mathbb{W}_w|} \sum_{v \in \mathbb{W}_w} x^T x - 2x^T \left(\frac{1}{|\mathbb{W}_w|} \sum_{v \in \mathbb{W}_w} v\right) - \\
&\qquad \frac{1}{|\mathbb{W}_w|} \sum_{v \in \mathbb{W}_w} M^T M + 2M^T \left(\frac{1}{|\mathbb{W}_w|} \sum_{v \in \mathbb{W}_w} v\right) \le \frac{k}{|\mathbb{W}_w|}\Bigg\} \\
= &\left\{x \in \mathbb{R}^n : x^T x - 2x^T M - M^T M + 2M^T M \le k/|\mathbb{W}_w|\right\} \\
= &\left\{x \in \mathbb{R}^n : x^T x - 2M^T x + M^T M \le k/|\mathbb{W}_w|\right\} \\
= &\left\{x \in \mathbb{R}^n : \|x - M\|^2 \le k/|\mathbb{W}_w|\right\} \\
= &\left\{x \in \mathbb{R}^n : \|x - M\| \le \sqrt{k/|\mathbb{W}_w|}\right\}.
\end{aligned}$$

So, the optimal value of

$$\min \left\{\sum_{v \in \mathbb{W}_w} \|w' - v\|^2 : w' \in \mathbb{W}_w\right\} \tag{4.4}$$

is $\sum_{v \in \mathbb{W}_w} \|M - v\|^2 + k$, where $k$ is the smallest value such that

$$\left\{x \in \mathbb{R}^n : \|x - M\| \le \sqrt{k/|\mathbb{W}_w|}\right\} \cap \mathbb{W}_w \neq \emptyset.$$

Since the first set is a ball around $M$ of radius $\sqrt{k/|\mathbb{W}_w|}$, we have from the definition of $\mathrm{proj}_{\mathbb{W}_w}(M)$ that the smallest value $k/|\mathbb{W}_w|$ with this property is $(M - \mathrm{proj}_{\mathbb{W}_w}(M))^T(M - \mathrm{proj}_{\mathbb{W}_w}(M))$, which guarantees

$$\mathrm{proj}_{\mathbb{W}_w}(M) \in \left\{x : \|x - M\| \le \|M - \mathrm{proj}_{\mathbb{W}_w}(M)\|\right\} \cap \mathbb{W}_w.$$

Hence, $\mathrm{proj}_{\mathbb{W}_w}(M)$ solves (4.4). ∎

Theorem 4.3 shows that an element of (4.1) can be calculated by projecting the center-of-mass onto $\mathbb{W}_w$, provided that $\alpha(v)$ is constant and $\mathbb{V}$ is finite. Both calculating the center-of-mass and projecting it onto $\mathbb{W}_w$ are $O(|\mathbb{V}|)$, which means the complexity of calculating $\hat{\mathbb{W}}$ in the Lloyd algorithm is $O(N|\mathbb{V}|)$. Unlike the Lloyd algorithm, the Maranzana heuristic was developed without any foreknowledge of $\gamma$ and $\beta$, and hence the construction of $\hat{\mathbb{P}}$ requires pairwise comparisons within each $\mathbb{V}_p$, which is $O(N|\mathbb{V}|^2)$. So, in the finite case when $\beta$ is constant and $\gamma$ is squared error, the Lloyd algorithm applied to the $N$-median problem has significantly lower complexity, a fact supported by the numerical results of the next section.

The conditions of Theorem 4.3 are limiting, and in general, the pairwise comparisons of the Maranzana algorithm are needed. In fact, translating an $N$-median problem into a VQ problem that naturally mirrors the framework presented in the VQ literature is not possible. This is because the similarity measures (edge weights) $\gamma(v_i, v_j)$ are not generally assumed to represent the geometry of $\mathbb{R}^n$ whereas the values $\rho(v_i, v_j)$ generally do. Our algebraic models easily demonstrate the relationship between the two problems, but to adapt a solution methodology for VQ design, like the Lloyd algorithm that uses the metric properties of $\mathbb{R}^n$, requires caution. If $\gamma$ does not allow the graph to be embedded in $\mathbb{R}^n$, then there is not a natural way to give the nodes coordinates. Without coordinates, the arithmetic needed to calculate the center-of-mass is not well defined, making the Lloyd algorithm useless for the general $N$-median problem.

Heuristics developed for the $N$-median problem are designed to work with or without any special structure, and it is simple to translate VQ problems into $N$-median problems and use any number of heuristics developed for the $N$-median problem, see [11] for a review of such methods. Because of this, the numerical experiments of the next section consider heuristics to the $N$-median problem and the Lloyd algorithm (appropriate structure is assumed). The most common and most tested heuristic for the $N$-median problem is vertex substitution (VS), see [14, 15, 16] for comparisons.

Vertex substitution was originally developed in 1968 by Teitz and Bart [17]. Each iteration of the method decides whether or not to swap a position in $\mathbb{P}$ with position not in $\mathbb{P}$. Variations differ in how they select the elements to swap. In 1983, Whitaker [19] developed an implementation known as fast interchange, which was later implemented in the Variable Neighborhood Search method of Hansen and Mladenovicć [7]. Both of these implementations begin by searching through $\mathbb{V}\backslash\mathbb{P}$ and testing whether swapping with an element of $\mathbb{P}$ would reduce the objective function. Whitaker's method performs the swap with the first profitable position found, whereas the Hansen and Mladenović implementation tests all possible swaps and performs the most profitable one. Interested readers are directed to [7] and [12] for complete descriptions.

# 5 Numerical Experiment

The discussions of Section 4 identify possible performance differences in the solution methods for the two problems, and in this section we numerically compare the performance of the methods presented in Section 4. To make sure that all problems can be solved with the Lloyd algorithm in addition to the heuristics for the $N$-median problem, we assume all examples are complete digraphs for which $\mathbb{P}' = \mathbb{V} \subseteq \mathbb{R}^3$, $|\mathbb{V}| < \infty$, $\beta(v) = 1$ and $\gamma(v_1, v_2) = \|v_1 - v_2\|^2$. This graph representation allows us to naturally interpret the problem of finding medians as a problem in VQ design, simply let $\alpha = \beta$ and $\rho = \gamma$.

Problems are identified by the tuple $(|\mathbb{V}|, N)$, so $(1000, 15)$ is an instance with 1000 nodes (vectors) and 15 medians (codewords). We randomly generated $\mathbb{V}$ with MATLAB 7.0 and considered all instances $(|\mathbb{V}|, N)$ in

$$\{100, 250, 500, 1000, 1200, 1500, 2000\} \times \{5, 10, 15, 20, 30\},$$

producing 35 different problems in the unit hypercube of $\mathbb{R}^3$. For each problem we randomly generated 30 different $N$-element subsets of $\mathbb{V}$ to use as starting points for each heuristic.

Each problem was modeled as the standard binary optimization problem in Appendix A and solved with CPLEX's network simplex algorithm. This allowed us to know a global solution for problems with $|\mathbb{V}| \leq 1000$ (larger instances were beyond this technique) together with the time needed to calculate it. Other CPLEX options were considered, but the network simplex method consistently outperformed the other possibilities.

The discrete Lloyd algorithm dominates the other techniques with respect to speed, but as Figure 3 indicates, the solution quality is not as impressive as Hansen's approach, which is routinely within 5% of the global optimum. A natural question was whether or not we could harness the speed of the discrete Lloyd algorithm to seed the vertex substitution methods to improve run time. This experiment was conducted, and each problem with each starting point was solved by 6 heuristics: Maranzana's algorithm, the discrete Lloyd algorithm, Hansen's algorithm, Whitaker's algorithm, and both Hansen's and Whitaker's technique initialized with the solution from the discrete Lloyd algorithm. All implementations were written in MATLAB, and results are reported in terms of the mean and standard deviation of the objective value, number of iterations, and run times over the 30 solves for each problem.

The results for problem instances of size $|\mathbb{V}| = 500$ are shown in Table 1. A complete list of tables for all cases is found at `http://lagrange.math.trinity.edu/tumath/research/reports/misc/report102/`. The numbers in parentheses indicate the appropriate percentage of the global solution found by the network simplex algorithm. For example, a value of (1.10) in the Objective column indicates that the heuristic terminated with an objective value that was 110% of the global optimum and a value of (0.72) in the time column means the heuristic required 72% of the time needed to find the global solution.

As Theorem 4.3 indicates, the Maranzana and discrete Lloyd algorithm terminate with identical solutions, but our numerical results mirror the complexity analysis

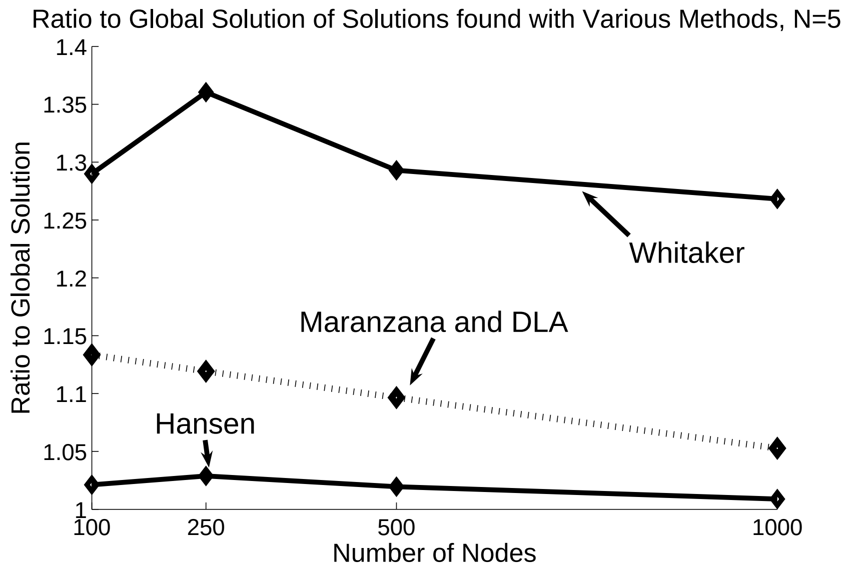


Figure 3: The average objective value as a ratio of the global solution versus $|\mathbb{V}|$ with $N = 5$. Each technique appears to improve its solution quality as $|\mathbb{V}|$ increases.

of the previous section and show that the discrete Lloyd algorithm has a significant computational advantage over the Maranzana algorithm, see Figures 4 and 5. Figure 5 is the same as Figure 4 except that it includes the solution time for the network simplex algorithm for $|\mathbb{V}| \leq 500$, and this curve appears exponential. The heuristics clearly dominate the global approach with respect to speed. For a few of the smaller networks the heuristics were slower than the network simplex algorithm, something we attribute to the fact that the heuristics were coded in MATLAB, which is interpreted.

Figure 6 shows an odd trend in the solution quality of the discrete Lloyd and Maranzana algorithms. Allowing $|\mathbb{V}|$ to remain constant, notice that the solution quality degrades as $N$ increases. The ratios of Whitaker and Hansen are nearly constant at 1.02 and 1.3, respectively. This indicates that vertex substitution is less sensitive to a change in $N$. Each technique took longer to converge as $N$ increased, although the change for the discrete Lloyd method was insignificant.

In general, Hansen's approach obtains excellent solutions but takes a long time to solve. Whitaker's approach is faster but produces solutions of lesser quality. The discrete Lloyd and Maranzana methods are even faster and produce solutions generally better than Whitaker's method. However, these techniques are sensitive to $N$, and it appears as though solution quality approaches that of Whitaker's approach as $N$ increases. In many cases initializing Whitaker's method with the solution from the Lloyd algorithm improved the solution quality, although the improved solution was still not as good as that from Hansen's technique. Initializing Hansen's technique

| | | | Objective | | Iterations | | Time | |
|---|---|---|---|---|---|---|---|---|
| $\lvert\mathbb{V}\rvert$ | $N$ | Method | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| | | Global | 501407.97 (—) | — | 114294 | — | 4282.88 (—) | — |
| | | MAR | 549777.60 (1.10) | 26591.23 | 4.13 | 1.25 | 2.15 (0.00) | 0.12 |
| | | DLA | 549777.60 (1.10) | 26591.23 | 4.13 | 1.25 | 0.03 (0.00) | 0.01 |
| 500 | 5 | HAN | 511191.39 (1.02) | 7201.95 | 9.87 | 2.11 | 31.22 (0.01) | 6.24 |
| | | WHIT | 648306.88 (1.29) | 50385.48 | 4.27 | 2.42 | 7.61 (0.00) | 2.62 |
| | | HAN / DLA | 511004.83 (1.02) | 8521.28 | 8.00 | 2.84 | 26.56 (0.01) | 10.64 |
| | | WHIT / DLA | 545965.52 (1.09) | 19918.92 | 1.33 | 0.61 | 5.34 (0.00) | 1.29 |
| | | Global | 274236.58 (—) | — | 30483 | — | 299.22 (—) | — |
| | | MAR | 311046.20 (1.13) | 16103.83 | 5.20 | 1.94 | 3.09 (0.01) | 0.74 |
| | | DLA | 311046.20 (1.13) | 16103.83 | 5.20 | 1.94 | 0.12 (0.00) | 0.07 |
| 500 | 10 | HAN | 278217.88 (1.01) | 3492.21 | 17.87 | 3.51 | 216.63 (0.72) | 90.27 |
| | | WHIT | 392236.36 (1.43) | 19055.32 | 7.47 | 3.41 | 34.08 (0.11) | 20.24 |
| | | HAN / DLA | 277250.25 (1.01) | 3842.16 | 13.10 | 4.17 | 161.88 (0.54) | 79.07 |
| | | WHIT / DLA | 309458.05 (1.13) | 14132.92 | 1.17 | 0.38 | 14.61 (0.05) | 6.13 |
| | | Global | 209176.79 (—) | — | 142286 | — | 7465.60 (—) | — |
| | | MAR | 241080.60 (1.15) | 11371.49 | 4.57 | 1.19 | 2.68 (0.00) | 1.01 |
| | | DLA | 241080.60 (1.15) | 11371.49 | 4.57 | 1.19 | 0.11 (0.00) | 0.07 |
| 500 | 15 | HAN | 213132.48 (1.02) | 1963.72 | 23.47 | 4.06 | 329.44 (0.04) | 171.43 |
| | | WHIT | 276778.90 (1.32) | 13252.56 | 12.40 | 3.91 | 55.78 (0.01) | 40.46 |
| | | HAN / DLA | 213871.82 (1.02) | 2478.79 | 15.73 | 4.60 | 234.54 (0.03) | 157.51 |
| | | WHIT / DLA | 238216.37 (1.14) | 9649.00 | 1.63 | 0.85 | 19.59 (0.00) | 14.52 |
| | | Global | 167475.09 (—) | — | 31609 | — | 3427.54 (—) | — |
| | | MAR | 198087.25 (1.18) | 6001.96 | 4.77 | 1.25 | 3.61 (0.00) | 0.81 |
| | | DLA | 198087.25 (1.18) | 6001.96 | 4.77 | 1.25 | 0.26 (0.00) | 0.11 |
| 500 | 20 | HAN | 171428.21 (1.02) | 1849.10 | 28.63 | 4.96 | 939.15 (0.27) | 172.05 |
| | | WHIT | 219462.79 (1.31) | 8785.06 | 16.80 | 5.37 | 184.54 (0.05) | 61.14 |
| | | HAN / DLA | 171983.64 (1.03) | 2228.37 | 20.37 | 5.01 | 706.38 (0.21) | 186.52 |
| | | WHIT / DLA | 193251.61 (1.15) | 4651.41 | 2.87 | 1.38 | 47.11 (0.01) | 19.43 |
| | | Global | 121038.81 (—) | — | 14912 | — | 795.63 (—) | — |
| | | MAR | 151675.27 (1.25) | 5472.31 | 5.17 | 1.23 | 3.32 (0.00) | 0.46 |
| | | DLA | 151675.27 (1.25) | 5472.31 | 5.17 | 1.23 | 0.37 (0.00) | 0.11 |
| 500 | 30 | HAN | 123840.27 (1.02) | 954.77 | 37.93 | 3.52 | 1942.48 (2.44) | 249.03 |
| | | WHIT | 157963.85 (1.31) | 4979.40 | 23.60 | 6.03 | 331.29 (0.42) | 108.69 |
| | | HAN / DLA | 124195.91 (1.03) | 1284.02 | 27.70 | 5.03 | 1414.40 (1.78) | 307.38 |
| | | WHIT / DLA | 144139.45 (1.19) | 3200.81 | 5.53 | 2.40 | 96.13 (0.12) | 32.92 |

Table 1: Data on problem instances with $|\mathbb{V}| = 500$ and varying values of $N$ and solution methods. The number of runs of each heuristic technique was 30.

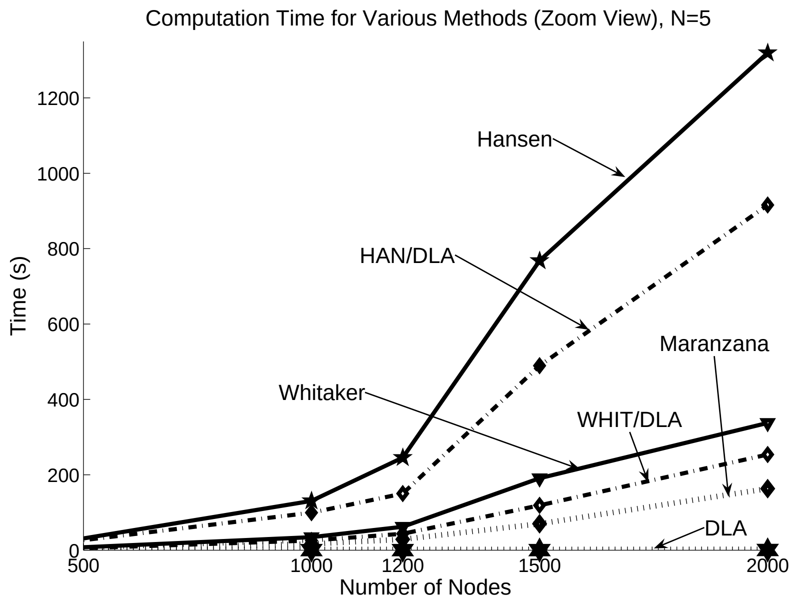


Figure 4: The computation time for the heuristics plotted versus $|\mathbb{V}|$.

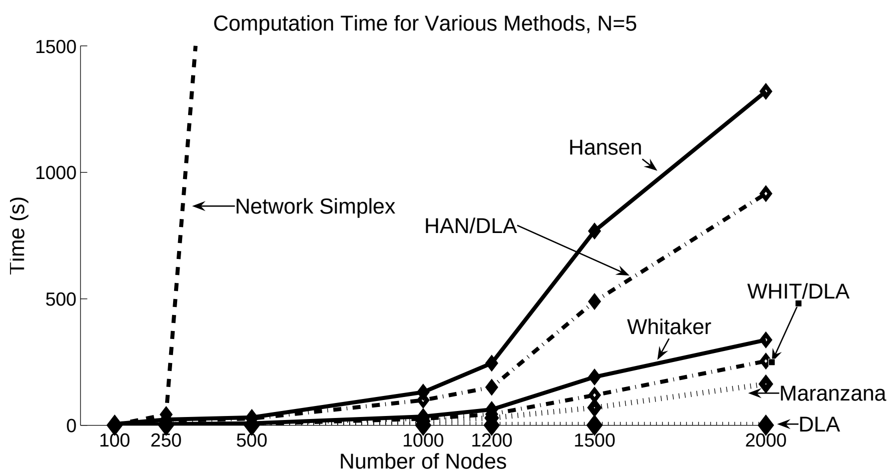


Figure 5: The computation time plotted versus the value of $|\mathbb{V}|$. The solve time for the network simplex algorithm with $|\mathbb{V}| = 1000$ was 675021.47sec $\approx$ 8days

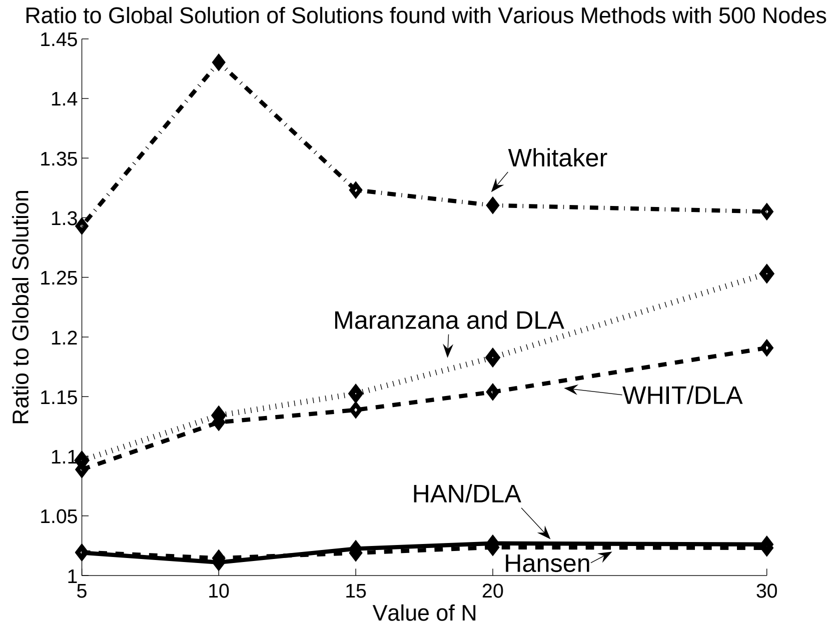


Figure 6: The objective value as a ratio of the global solution versus $N$ with $|\mathbb{V}| = 500$.

with the solution from the discrete Lloyd algorithm did not produce a remarkable change, which is expected since the solutions were generally close to optimal anyway.

In the future we hope to improve the solution quality of the discrete Lloyd algorithm without sacrificing its favorable speed. In particular, we hope to be able to initialize the algorithm so that it converges to a near optimal solution. Also, the numerical results clearly demonstrate that Hansen's approach produces quality solutions in a fraction of the time it takes to calculate the global solution, and Theorem 3.1 allows us to use this technique to solve VQ problems. We suspect that the numerous applications of VQ design could benefit from this approach.


**Acknowledgment**

The authors thank Paul Uhlig for his suggestion on the technique of proof for Theorem 4.3.


# References


[1] S. Durocher and D. Kirkpatrick. The projection median of a set points in $\mathbb{R}^2$. In *Proceedings of the Seventeenth Canadian Conference on Computational Geometry*, pages 17:46–50, 2005.

[2] M. R. Garey and D. S. Johnson. *Computers and Intractibility: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Co., San Francisco, 1979.

[3] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 1992.

[4] H.J. Greenberg. *Mathematical Programming Glossary*. World Wide Web, `http://glossary.computing.society.informs.org/`, 1996–2006. Edited by the INFORMS Computing Society.

[5] H.J. Greenberg and S. Robinson. Personal communications, 2006.

[6] S. L. Hakimi. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research*, 13(3):462–475, 1965.

[7] P. Hansen and N. Mladenović. Variable neighborhood search for the $p$-median. *Location Science*, 5(4):207–226, 1997.

[8] O. Kariv and S. L. Hakimi. An algorithmic approach to network location problems. II. The $p$-medians. *SIAM Journal on Applied Mathematics*, 37(3):539–560, 1979.

[9] S. P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inform. Theory*, 28:129–137, 1982.

[10] F. E. Maranzana. On the location of supply points to minimize transport costs. *Operational Research Quarterly*, 15(3):261–270, 1964.

[11] J. Reese. Solution methods for the $p$-median problem: An annotated bibliography. *Networks*, 48(3):125–142, 2006.

[12] M. G. C. Resende and R. F. Werneck. On the implementation of a swap-based local search procedure for the $p$-median problem. In R. Ladner, editor, *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 119–127, Philadelphia, 2003. SIAM.

[13] C. ReVelle and R. Swain. Central facilities location. *Geographical Analysis*, 2:30–42, 1970.

[14] K. E. Rosing, E. L. Hillsman, and H. Rosing-Vogelaar. A note comparing optimal and heuristic solutions to the $p$-median problem. *Geographical Analysis*, 11(1):86–89, 1979.

[15] K. E. Rosing, E. L. Hillsman, and H. Rosing-Vogelaar. The robustness of two common heuristics for the $p$-median problem. *Environment and Planning A*, 11:373–380, 1979.

[16] G. Rushton and J. A. Kohler. ALLOC: Heuristic solutions to multi-facility location problems on a graph. In G. Rushton, M. Goodchild, and L. Ostresh, editors, *Computer Programs for Location-Allocation Problems*, pages 163–187. Monograph No. 6, Department of Geography, University of Iowa, Iowa City, 1973.

[17] M. B. Teitz and P. Bart. Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research*, 16(5):955–961, 1968.

[18] G. Wesolowsky. The Weber problem: History and perspectives. *Location Science*, 1(1):5 – 23, 1993.

[19] R. A. Whitaker. A fast algorithm for the greedy interchange of large-scale clustering and median location problems. *INFOR*, 21(2):95–108, 1983.

# A Binary Representation of the $N$-Median Problem

The following binary formulation of the $N$-median problem from [13] was used in our numerical experiments. Let $\xi_{ij}$ be such that

$$\xi_{ij} = \begin{cases} 1 & \text{if vertex } v_i \text{ is allocated to vertex } v_j \\ 0 & \text{otherwise.} \end{cases}$$

With this notation, the math program is

$$\begin{array}{rrcl} \min & Z & = & \sum_{ij} \beta(v_i)\gamma(v_i, v_j)\xi_{ij} \\ \text{subject to} & \sum_j \xi_{ij} & = & 1, \text{ for } i = 1, \ldots, n, \\ & \sum_j \xi_{jj} & = & N, \\ & \xi_{jj} & \geq & \xi_{ij}, \quad \forall i, j = 1, \ldots, n, \\ & \xi_{ij} & \in & \{0, 1\}. \end{array}$$

The first constraint ensures that each vertex is allocated to one and only one element in the $N$-element subset. The second constraint guarantees that there are $N$ vertices allocated to themselves, which forces the cardinality of the $N$-median subset to be $N$. Although this is not a requirement of the general $N$-median problem we presented, it is appropriate for our numerical work in which $\gamma(v_i, v_j) = \|v_i - v_j\|^2$. The third constraint states that vertices cannot be allocated to non medians. The solution is $\{v_j \mid \xi_{jj} = 1\}$.